HIERARCHICAL MULTILABEL CLASSIFICATION

Pooja D. Bankar[1] S.S.Sane[2]

[1]M.E. Student [2]HOD & Vice Principal
[1,2]Department of Computer Engineering
[1,2] K. K. W. I. E. E. R., Nashik
Savitribai Phule Pune University, Maharashtra, India
[1]poojabnkr@gmail.com

ABSTRACT

Multi-label classification assigns multiple labels for every instance. If labels are ordered in predefined structure called hierarchical multi-label classification (HMC). HMC is carried out using two approaches such as top down (or local) and one shot (or global). Top down approach splits the problem into sub-problems while one shot approach considers the problem as a whole. One shot approach faces a problem of explodation with huge datasets. Chain Path Evaluation (CPE) method provides solution to address this issue using either Direct Acyclic Graph or tree data structures. The method combines predictions of local classifiers based on the node's relationships and position in the hierarchy. Probability based pruning is used by CPE and thus found to be superior when compared with other methods with respect to accuracy. This work aims at implementation of CPE and extend it further to remove label duplication. Experimental results show that removal of label duplication results in better accuracy for both MLNP and NMLNP tree based data structures.

Keywords: *chain classifiers, hierarchical classification, multi-label classification*

1. INTRODUCTION

Traditionally classification task deals with problems where only single label is used, also called as multiclass classification problem. Some classification problems can have multiple labels and more complex. There are two types of approaches for classification such as multilabel classification and multidimensional classification. Classification of binary classes is called multilabel classification and classes having multiple class values are called multidimensional classification.

Consider a multilabel dataset D having N instances, every example e is correlated with set Y of labels where Y⊆ L and L is the set of classes. Hierarchical multilabel classification [1] is that when the labels are ordered in predefined structure. There are two hierarchical structures such as tree or a DAG. In tree structure node has maximum one parent and in DAG structure a node can have more than one parent node. Fig. 1 shows tree structure and Fig. 2 shows DAG structure.

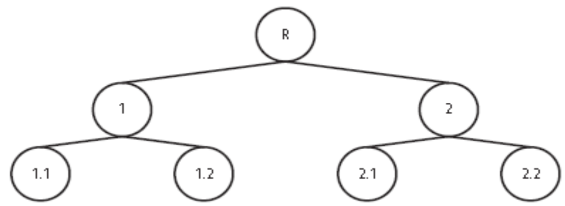

Fig 1. Tree structure

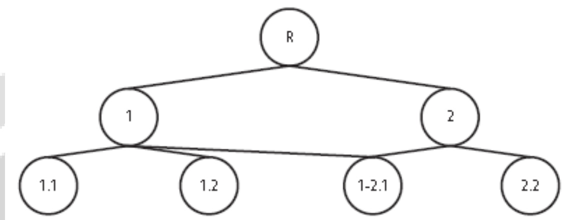

Fig 2. DAG structure

HMC includes two approaches such as top down (or local) and one shot (or global). Top down approaches split the problem into sub-problems while one shot approaches consider the problem as a whole. Local approach faces the problem with huge datasets. Therefore, the novel HMC approach such as Chain Path Evaluation (CPE) is proposed that belongs to the local approach and works efficiently with huge datasets. The prediction of parent node is done by adding extra attribute to the instances of each node to show the relationship between classes. It combines the weighting scheme for the prediction of more general classes. By scoring all the paths it selects the best one. CPE contains pruning phase that discards the branches having less probability.

CPE predicts multiple paths for DAG hierarchies and single path for tree hierarchies. CPE performs better than other methods.

2. RELATED WORK

The same node appears in various fields or can belong to two or more classes are a multilabel classification such as, text categorization, where one document belongs to many topics.

When organised classes are in the form of hierarchy then it is a hierarchical multilabel classification.

When the predefined structure of ordered labels in case of tree or DAG structure then it is called as hierarchical multilabel classification. The structure of class is shown using IS-A relations. The relations are transitive and asymmetric.

There are two types of predictions [1] in HMC such as mandatory leaf node prediction (MNLP) and non-mandatory leaf node prediction (NMNLP). In MNLP path ends at leaf node and in NMNLP path may end at non-leaf node.

HMC methods are grouped into three categories namely flat, local and global.

Flat classification approach does only prediction of leaf nodes and ignores the class hierarchy. It behaves like traditional multilabel classification algorithm. The serious disadvantage of this approach is that it requires to generate classifier on large number of leaf nodes and it doesn't take advantage of the information of class hierarchy.

The global approach is a single decision model and predicts each node of the hierarchy. While single run of classification algorithm this model considers whole class hierarchy. The disadvantage of this approach is that they are very complex and requires more time when data size increases. The global HMC method is presented by Vens et al. [2], which is a decision tree algorithm based on Predictive Clustering Trees (PCT).

Local classifiers have three standard ways to use local information: 1) a local classifier per level (LCL), 2) a local classifier per parent node (LCPN), 3) a local classifier per node (LCN) and. All these three types differ in training phase but similar top down approach in testing phase. The disadvantage of this top down approach is that the error which occurs in an upper level propagates down to the hierarchy.

Local classifier per node (LCN) approach trains classifier for each node of the class hierarchy called binary classifier and it generates classifier without considering the root node. Bi and kwok [3,4] proposed HIROM, method that uses the local prediction to search for multilabel classification. Prediction rule is derived using Bayesian decision theory. The disadvantage of this approach is that it does not maximize the performance measure.

Local classifier per level (LCL) approach trains classifier on each node which is a multi-class classifier. The method that trains classifier for each level is proposed by a Cerri et al. [5]. It is used as baseline comparison approach.

Local classifier per parent node (LCPN) approach trains classifier for class hierarchy which is a multi class classifier for each parent node. Top down prediction approach is used during testing phase. LCPN algorithm is proposed by Silla et al. [6] with two training methods. The best features are selected to train the classifiers in first method; the best classifier and subset of features are selected in second method, which improves the classification performance. It has fewer classifiers than LCN. The disadvantage of this approach is that it may generate blocking problem and can cause to inconsistency.

The LCPN is proposed by Hernandez et al. [7] for tree structured datasets. In the classification method, LCPN classifies a new instance having local classifier at each node, from the root node to a leaf node for each path a score is obtain by containing the results of all nodes. Two fusion rules such as product rule and sum rule are used to obtain these outputs. Finally, it gives the path having best score.

Unlike other local approaches CPE takes the information given by the class hierarchy with consideration of the predictions of parent nodes. CPE also deals with DAG structure and gives non-mandatory leaf node predictions.

3. SYSTEM ARCHITECTURE

Chained path evaluation

Hernandez et al. [7] work is extended to the chain path evaluation. To include the relations between the labels, classifiers are trained, mostly the parent nodes for accurate prediction. CPE evaluates the score for each path using merging rule and return one with best score. Before evaluation of final prediction, it uses pruning for eliminating uncertain labels.

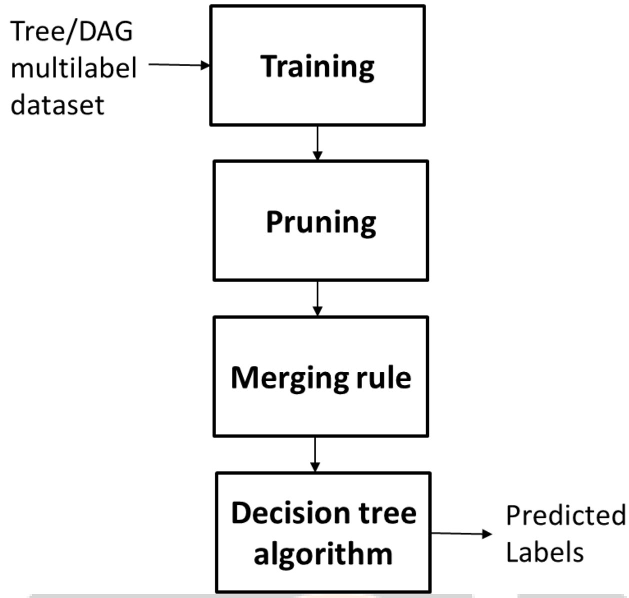

Figure 1: Block diagram of CPE-DT

A. Training

Consider D be a training dataset with N examples, e_e = (Xe, Ye), where Xe is a feature vector and Ye ⊆ L, L is the set of possible classes or labels, L= $\{Y_1, Y_{2, \ldots\ldots} Y_{|L|}\}$. Ye is small set than the L.

Training set with feature vector and number of labels is taken as input to the training phase. Multi class classifier is trained for each non-leaf node using LCPN approach called as base classifier. Base classifier can be any classifier as it returns probability value of predicted class. Positive and negative training sets are calculated for each class. Positive training set contains child of that class and negative training set contain siblings of that class. In this phase, balanced training set is formed.

B. Pruning

Pruning is used for non-mandatory leaf node prediction. Sometimes the information related to the lower level instances is not required. Pruning can be done in two ways: top down and bottom up way. This framework uses top down approach of pruning. The paths in the tree or DAG are pruned using pruning.

C. Merging rule

After pruning the hierarchy, merging rule sets the score to each path from root to the leaf node. Error in the upper hierarchy level also misclassifies to the lower hierarchy i.e. wrong path from beginning gives prediction of whole path.

To achieve this rule, weight of each node and its level both are calculated. level(y_i) is the level at which the node y_i is placed in hierarchy.

$$\text{level}(yi) = 1 + \frac{\sum_{j=1}^{m} level(pa(yi)j)}{|pa(yi)|} \qquad (1)$$

For tree structure, one is added to the level of its parent, and for DAG structure, one is added to the mean of the level of its parent. w(yi) is the weight of the node.

w(yi)= $1 - \frac{level(yi)}{maxLevel+1}$ (2)

Where, maxLevel is the longest path in the hierarchy.

Equation (3) describes the merging rule which is the weighted sum of the logarithms of the probabilities predicted on nodes along the path.

Score=$\sum_{i=1}^{p} w(yi) * (p(yi|xe, pa(yi)))$ (3)

Where, p= number of nodes in the path,

yi= ith node in the path, and

P(yi=1|Xe,pa(yi))= probability of node yi predicted as true by local classifier.

In Dag structure, there might be multiple paths but in this case the paths that end to the leaf node are returned. The merging rule predicts single path which is converge to a leaf node by combining all local classifiers. The path having best score is returned as final prediction.

D. Decision tree algorithm

The use of this algorithm is to remove the duplication of labels having best score. The best score is used to compare the duplicate labels. This algorithm works as follows.

1. Select one label.
2. Find that selected label having maximum score.
3. Compare label having maximum score with labels having less than or equal score.
4. Remove the duplicate labels.
5. Finally, we get set of predicted labels

Experimental results

There are two types of genomic datasets, one is the MLNP dataset in which the instances of hierarchy reach a leaf node and second is NMLNP dataset in which the instances of hierarchy do not reach a leaf node.

To evaluate the CPE five datasets are used. In which three datasets are tree structured hierarchy and two are DAG structured hierarchy. funCat (FUN)[8] is used for tree structure and Gene ontology (GO)[9] is used for DAG structure. FunCat is a tree-structured class hierarchy. In the second version of the data sets, the genes are annotated with terms from the Gene Ontology (GO), which forms a directed acyclic graph instead of a tree: each term can have multiple parents (we use GO's "is-a" relationship between terms).

The below table 9.3 shows labels predicted by CPE-DT method for Tree and DAG datasets.

Table 9.3: Predicted labels of given datasets

Dataset	Predicted labels
(a)MLNP datasets	
cellcycle_FUN	13
church_FUN	08
derisi_FUN	09
seq_GO	21
spo_GO	14
(b)NMLNP datasets	
cellcycle_FUN	11
church_FUN	08
derisi_FUN	10
seq_GO	17
spo_GO	11

Conclusion

The existing system uses CPE method to classify hierarchical multilabel datasets. In this the pruning method is used to remove the non mandatory leaf nodes and gives the path having best score. Proposed work uses the same algorithm with extension to remove label duplication. Removal of label duplication results in better accuracy for both MLNP and NMLNP tree based data structures as compared with other methods.

REFERENCES

[1]. C.N. Silla Jr., A.A. Freitas,"A survey of hierarchical classification across different application domains," Data Min. Knowl. Discov. 22 (April 2010) 31-72.

[2]. C. Vens, J. Struyf, L. Schietgat, S. Dzeroski, H. Blockeel," Decision trees for hierarchical multi-label classification," Mach. Learn. 73(2) (2008) 185214.

[3]. W. Bi, J.T. Kwok,"Multi-label classification on tree-and DAG-structured hierarchies," in: 28th International Conference on Machine Learning (ICML-11), 2011, pp.1724.

[4]. W.. Bi, J.T. Kwok,"Hierarchical multilabel classification with minimum Bayesrisk," in: Data Mining (ICDM), 2012 IEEE 12th International Conference, December 2012.

[5]. Cerri, R., Barros, R. C., de Carvalho, A.C.P.L.F.,"Hierarchical multilabel classification using local neural networks," J. Comput. System Sci. 1, 1-18(2013.

[6]. Silla Jr., C. N., Freitas, A.A.:" A novel top-down approaches for hierarchical classification and their application to automatic music genre classification," in: IEEE Inter. Conf. on Systems, Man, and Cybernetics, pp. 3499-3504(2009).

[7]. Hernandez, J.N., Sucar, L.E., Morales, E.F,"A hybrid global-local approach for hierarchical classification," In: Boonthum-Denecke, C., Youngblood, G.M. (eds.) Proceedings FLAIRS Conference, pp. 432437. AAAI Press (2013).

[8]. A. Ruepp, A. Zollner, D. Maier, K. Albermann, J. Hani, M. Mokrejs, I. Tetko, U. Guldener, G. Mannhaupt, M. Munsterkotter, H.W. Mewes, "The FunCat, afunctional annotation scheme for systematic classification of proteins from whole genomes", Nucleic Acids Res. 32(18) (2004) 5539-5545.

[9]. M. Ashburner, C.A. Ball, J.A. Blake, "Gene ontology: tool for the unification of biology",Nat. Genet. 25 (2000) 25-29.